

Question 6: Free the Friulian three

Authors: Dick Hudson and Harold Somers

Friulian is sometimes referred to as a dialect of Italian, but it is sufficiently different to be classed as a separate language, spoken by around 600,000 people in Northeast Italy. Study the following Friulian phrases, then fill in the blanks in your answer book.

my house	<i>la mê cjase</i>
the houses	<i>lis cjases</i>
the bicycle	<i>la biciclete</i>
your bicycles	<i>lis tôs bicicletes</i>
our tickets	<i>i nestris tagongs</i>
my job	<i>il gno lavôr</i>
my sister	<i>mê sôr</i>
our brother	<i>nestri fradi</i>
your cousin	<i>to cusin</i>
my sisters	<i>lis mês sôrs</i>
our Irish grandmother	<i>la nestre ave irlandese</i>
my cousins	<i>i miei cusins</i>
our Irish cousin	<i>i nestri cusin irlandês</i>

- | | |
|---------------------|----------------------|
| (a) our house | (f) your sister |
| (b) the ticket | (g) our cousins |
| (c) my bicycle | (h) our jobs |
| (d) your brother | (i) our grandmothers |
| (e) my Irish sister | (j) my brother |

Question B: Georgia's always on my mind

20 points

Tbilisi is the capital of Georgia, a country in the Caucasus region of Eurasia. On the next page is a list of (some of) the stations on its subway (underground) system, written in Georgian, together with the equivalent English names of the stations, though not in the same order. While most of the names are simply transliterated, a few of them are rather translated and so do not match exactly. Note that J is pronounced like the 'y' in 'yes', and that the following sequences are regarded as single letters: DZ, SH, TS.



The Tbilisi metro system – don't worry if you can't read the names on this map, they are reproduced on the next page.

1	ვარკეთილი	A	Aragveli
2	სამგორი	B	Avlabari
3	ისანი	C	Didube
4	არაგველი	D	Gotsiridze
5	ავლაბარი	E	Guramishvili
6	სადგურის მოედანი	F	Isani
7	რუსთაველი	G	Marjanishvili
8	მარჯანიშვილი	H	Medical University
9	ნაძალადევი	I	Nadzaladevi
10	დიდუბე	J	Rustaveli
11	გურამიშვილი	K	Samgori
12	სარაჯიშვილი	L	Sarajishvili
13	ტექნიკური უნივერსიტეტი	M	Station Square
14	სამედიცინო უნივერსიტეტი	N	Technical University
15	გოცირიძე	O	Varketili

B1. Match up the station names 1-15 with their equivalents A-O.[15 points]

B2. How is the station name translated here as 'Station Square' pronounced in Georgian? [3 points]

B3. How is the Georgian word for 'Medical' pronounced? [2 points]

Question B Solution

B1.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
O	K	F	A	B	M	J	G	I	C	E	L	N	H	D

B2. ‘Station Square’ is pronounced **sadguris moedani**

B3. ‘Medical’ is pronounced **sameditsino**

Explanation

Although this may look complex because of the unfamiliar alphabet, it is actually quite a simple decoding question, once you realise that Georgian is a straightforward alphabet with consonants and vowels much like ours.. There are several “entry points” to the code, any one of which will get you started, after which you just have to patiently fill in the letters. One entry point would be the fact that there are two stations with “University” in their name, and two stations ending in უნივერსიტეტი (two of only three with two-word names). So you might reasonably try to match up the letters. The Georgian is actually *universiteti*, but the first part of the word will give you 6 or 7 letters. Another “way in” is to look for similar names such as three ending in *-ishvili* matching up with იშვილი. You could also count letters: *Isani* is the only 5-letter name, *Didube* the only 6-letter name. Or you could use the picture, the title of which, you might guess, says *Tblisis Metro*. How do you know which university is which? One is *sameditsino*, the other *texnikali*. To be fair, the *x* (pronounced like Irish/German/Welsh *ch*) doesn’t appear anywhere else. But it’s not hard to guess is it?

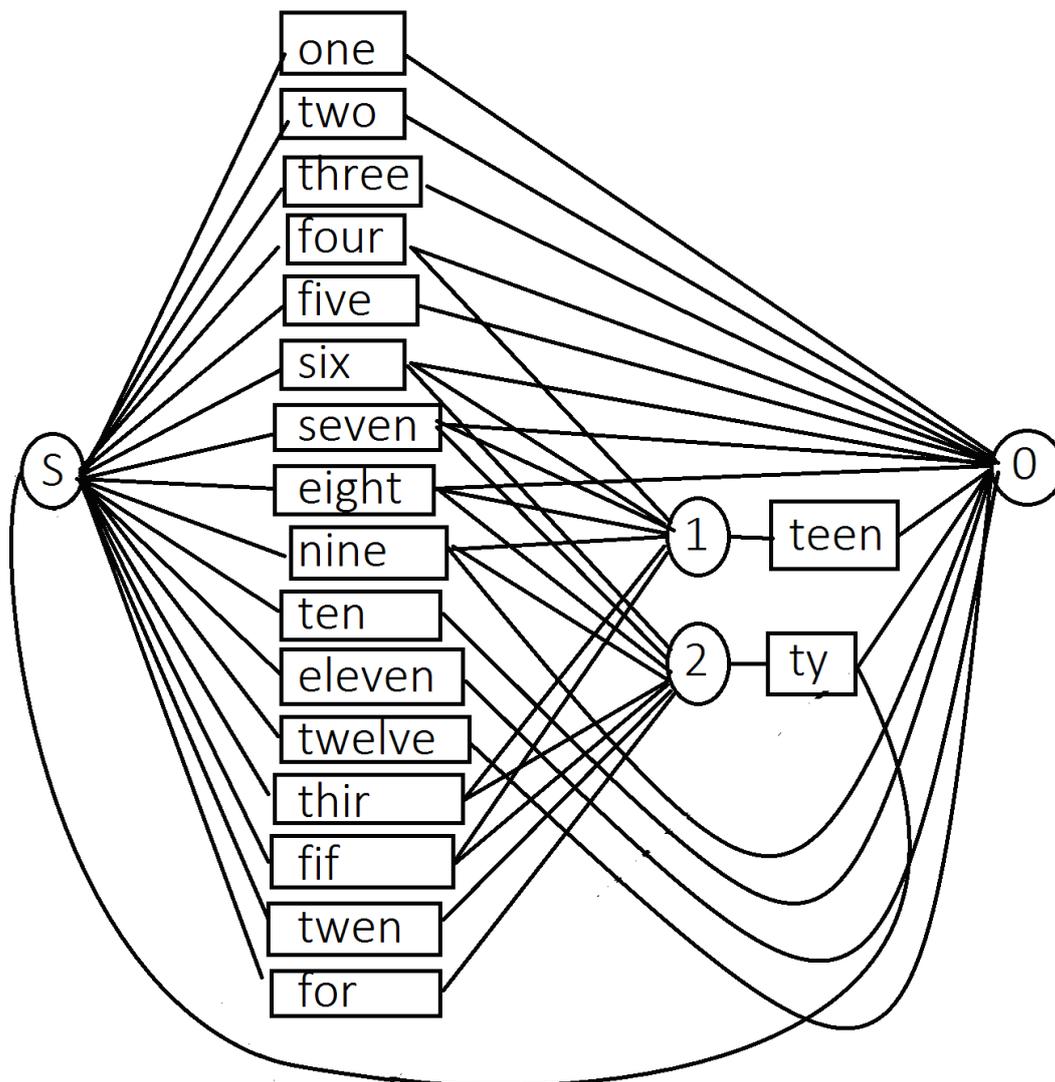
Here is the Georgian alphabet: not all the letters appeared in the data

ა ბ გ დ ე ვ ზ თ ი კ ლ მ ნ ო პ ჟ რ ს ტ უ ფ ქ ღ ყ შ ჩ ც ძ ჳ ჭ ხ ჯ
a b g d e v z t i k l m n o p zh r s t u p x gh q sh ch ts dz ts’ch kh j
Actually there are five more letters which are not used in modern Georgian.

Question C: Transitions

20 points

The diagram on the next page represents a “transition network”. The circles represent “states” while the boxes represent sequences of letters that can be “generated” from any given state, as indicated by the lines (the “transitions”). The aim is to start at “S” and get to the end state “0”. For some boxes there is a choice of transition. The lines are directional (it would have been even more messy to add the arrowheads), so note that you can only enter a state from the front (left). So you can go from “ty” to “S”, but not the other direction. As you can see, by following the paths you can generate the names of some numbers in English.



The above diagram is already quite messy, and it can be represented more neatly by a set of rules as below. Each rule is identified (in square brackets) but this is ONLY for ease of reference in answering the questions. Apart from that each rule consists of a state (the symbol before the “:”), a sequence of letters, and then, after the arrow (“→”) a list of states to which you can then move. Starting at position “S”, you generate the text

indicated, and then continue to any ONE of the rules whose start state is listed after the arrow. State “0” is a special case meaning “finish”.

[a]	S: one → 0	[j]	S: ten → 0
[b]	S: two → 0	[k]	S: eleven → 0
[c]	S: three → 0	[l]	S: twelve → 0
[d]	S: four → 0,1	[m]	S: thir → 1,2
[e]	S: five → 0	[n]	S: fif → 1,2
[f]	S: six → 0,1,2	[o]	S: twen → 2
[g]	S: seven → 0,1,2	[p]	S: for → 2
[h]	S: eight → 0,1,2	[q]	1: teen → 0
[i]	S: nine → 0,1,2	[r]	2: ty → S,0

So for example, starting at S we can generate “fourteen” by taking rule [d] to state 1, then rule [q] to finish. We cannot generate “twelveteen” because rule [l] only allows one way to progress, namely to finish.

C1. Write out the sequence of rules and states followed to generate the following words: For example, for “fourteen” write “d 1 q 0”. [12 points]

- (a) sixteen
- (b) ninetythree
- (c) twentyeight
- (d) fifteen

C2. The network above “overgenerates”, that is, it allows us to create sequences which are not valid number names. Indicate whether each of the following words can be generated by the network or not. [4 points]

- (a) oneten
- (b) fiftytwelve
- (c) sixteensix
- (d) twentyseventeen
- (e) fortythirty
- (f) eleventythree
- (g) fifty

C3. The above network wrongly generates a misspelling in the case of “eighteen” and “eighty...”. Suggest a simple fix for this (i.e. a change to one of the existing rules and an additional rule).

[4 points]

Question C Solution

C1. (a) (S) f 1 q 0

(b) (S) i 2 r S c 0

(c) (S) o 2 r S h 0

(d)(S) n 1 q 0

C2. Put a tick (✓) in the box if the network generates this string,

(a)	(b)	(c)	(d)	(e)	(f)	(g)
	✓		✓	✓		

C3. Change rule h as follows: S: eight \rightarrow 0

Additional rule S : eigh \rightarrow 1,2

(Some other fixes are possible, but this is the most elegant and obvious)

Explanation

Transition networks are a very simple tool in computational linguistics and simultaneously serve two purposes. One, seen here, is to capture the way a text string can be 'generated', i.e. start from S and follow the rules. But the network can also be used for 'parsing' or analyzing a string to see if it's legal. Actually that is what C2 asked *you* to do. What you did is probably similar to what a computer program would do. For example, to analyse 'sixteen', look for a rule that starts from state S and generates any of the first few characters: yes, [f] generates 'six'. Rule [f] says you can end (0) or go to state 1 or 2. Which rules can start from either of those states? Rules [q] and [r]. Do either of those rules lead you on to the next part of the string, 'teen'? Yes, rule [q]. And, crucially, does that rule then lead you to the exit state? Yes.

Let's try the same thing with C2e 'fortythyirty'. We want the network to fail for this. Let's see if it does. First we look for a rule that starts from state S and generates any of the first few characters: only rule [p] works. Note that rule [d] requires 'four', not 'for'. Rule [p] can only lead to one state, 2. And only one rule, [r] can lead off from that state. Rule [r] is looking for 'ty', which we do have. So far so good. Where can you go from rule [r]? Back to the start S, or finish (0). Can we generate 'thirty' from S? Yes we can (you should really trace the steps, but they are similar to 'forty': rule [m] then rule [r]. So 'fortythyirty' is legal, even though we don't want it to be.

Question D: Basquing in the sun

25 points

Basque is a language spoken by around 700,000 people in an area straddling France and Spain, in the western Pyrenees. Linguistically, Basque is unrelated to the other languages of Europe and indeed, is a “language isolate”, unrelated to any other known language.

Study the following sentences, shown in Basque together with their translations in English.

<i>Emakumeak gizona ikusi du.</i>	The woman has seen the man.
<i>Zuk umea ikusi duzu kalean.</i>	You have seen the child in the street.
<i>Non ikusi duzu umea?</i>	Where have you seen the child?
<i>Nork ikusi du umea kalean?</i>	Who has seen the child in the street?
<i>Umeak ez du gizona ikusi.</i>	The child has not seen the man.
<i>Emakumeak liburua irakurri du.</i>	The woman has read the book.
<i>Umea etorri da.</i>	The child has come/arrived.
<i>Umea kalean erori da.</i>	The child has fallen in the street.
<i>Igela agertu da.</i>	The frog has appeared.

D1. In your answerbook, translate (a)-(d) into Basque and (e)-(f) into English: [16 points]

- (a) The woman has come.
- (b) The man has seen the woman.
- (c) The book has not arrived.
- (d) The street has appeared in the book.
- (e) *Nork ikusi du gizona?*
- (f) *Igela kalean agertu da.*

Basque is a free word-order language. This means that all the phrases shown in Set A below translate into English as ‘The child has fallen in the street.’ (In English different word order would change the meaning, compare ‘The man bit the dog’ and ‘The dog bit the man’, but not so in Basque.)

SET A

Kalean umea erori da.
Umea erori da kalean.
Kalean erori da umea.
Erori da umea kalean.
Erori da kalean umea.

By contrast, all the phrases in Set B below are considered ungrammatical. That is, a native speaker of Basque would consider these structures unacceptable. Following convention, we mark ungrammatical sentences with an asterisk *.

SET B

- * *Erori umea da kolean.*
- * *Erori kolean da umea.*
- * *Erori umea kolean da.*

D2. In your answerbook, explain briefly the restriction which applies to the word order options available in Basque. [3 points]

Now consider the following additional data.

Gizonak umeari liburua eman dio.

The man has given the book to the child.

Irakasleak umeari liburua irakurrarazi dio.

The teacher has made the child read the book.

D3. Translate (a) into English, and (b) into Basque. [6 points]

(a) *Emakumeak umeari etorrarazi dio.*

(b) The child has given the frog to the woman.

Question D solution

- D1.** (a) Emakumea etorri da
(b) Gizonak emakumea ikusi du
(c) Liburua ez etorri da
(d) Kalea liburuan agertu da
(e) Who has seen the man?
(f) The frog has appeared in the street

D2. *Erori* can't be separated from *da*. More generally, the word order restriction is that a MAIN VERB and its following associated AUXILIARY VERB have to appear adjacent to each other and in that order.

- D3.** (a) The woman has made the child come
(b) Umeak emakumeari igela eman dio

Explanation

Basque nouns carry endings showing their function (what linguists call “case”) in the sentence. But unlike more familiar (Indo-European) languages, the cases are not exactly subject/object. Basque is a so-called “ergative” language, which means that the case corresponding to subject in a transitive sentence (where there is also an object) is not used for what would be the subject of an intransitive sentence (with no object): in that case the marking is the same as the object. The best way to understand it is to think of sentences like *John opens the door* and *the door opens*: in familiar languages (that show case marking like German, Latin, Greek and so on) – door would be object in the first, subject in the second. But in an ergative language it is the same case in both sentences, which has some logic when you think that the door plays the same role in both sentences (it gets opened). Linguists call this role “absolute”, while the “doer” or agent is called “ergative”. And quite simply, the nouns in the ergative case add a *-k*. There is also another case-marker in the first set of example sentences, which we can call locative *-n*. The endings occur on pronouns (*zuk, nork, non*) as well as nouns.

The verbs in the data mostly end in *-i*, but there is another word that needs explaining: *da, du, duzu*. This is an auxiliary verb (like *do* in English negatives, questions and past tenses) and takes the form *du* when there is an ergative (ending in *-k*), *da* when there is not, and *duzu* when ‘you’ is the ergative.

Word order is more or less free, so you could present the answers to D1 in any order, with the exception of *erori* which can't be separated from *da*.

Question E: Making a mark in Choctaw

20 points

Choctaw is a native American language spoken in Oklahoma and Mississippi by around 10,000 people.

Here are some Choctaw sentences with their English translations.

<i>Baliililitok.</i>	I ran.
<i>Baliilitok.</i>	He/she ran.
<i>Baliililih.</i>	I have run / I am running.
<i>Johnat niyah.</i>	John is fat.
<i>Pamat kayyah.</i>	Pam is pregnant.
<i>Baliilih.</i>	He/she has run <i>or</i> He/she is running.
<i>Chikayyah.</i>	You are pregnant.
<i>Saniyah.</i>	I am fat.
<i>Hattakat chaahah.</i>	The man is tall.
<i>Hattakat taloowatok.</i>	The man sang.
<i>Hattak chaahah piisalitok.</i>	I saw the tall man.
<i>Hattak chaahah ishpiisatok.</i>	You saw the tall man.

E1. Translate the following into Choctaw:

[10 points]

- (a) He/she saw John.
- (b) I sang.
- (c) You are singing.
- (d) I am pregnant.
- (e) He/she saw the fat man.

Here are a few more sentences or phrases in Choctaw:

<i>Sayyit sabashah.</i>	My leg is cut.
<i>Sapiisatok.</i>	He/she saw me.
<i>Issapiisatok.</i>	You saw me.
<i>amofi</i>	my dog
<i>Ofit sayyi sakopolitok.</i>	The dog bit my leg.
<i>Ofit amofi kopolitok.</i>	The dog bit my dog.
<i>Chimanolilitok.</i>	I told you.
<i>Amanolitok.</i>	He/she told me.
<i>Sashkit hattak piisatok.</i>	My mother saw the man.

E2. Translate into English:

[10 points]

- (a) *Chiyvit chibashah.*
- (b) *chimofi*
- (c) *Amofi ishpiisatok.*
- (d) *Chishki piisalitok.*
- (e) *Amofit chiyyi chikopolitok.*

Question E Solution

- E1.** (a) John piisatok.
(b) Taloowalitok.
(c) Ishtaloowah.
(d) Sakayyah.
(e) Hattak niyah piisatok.

- E2.** (a) Your leg is cut.
(b) your dog
(c) You saw my dog.
(d) I saw your mother.
(e) My dog bit your leg.

Explanation

Word-order: the verb comes last.

Pronouns are unusual as the verb markings are sufficient to convey subject, object etc.

Direct object nouns are unmarked, but subject nouns take the suffix *-at* if the noun ends in a consonant or *-t* if the noun ends in a vowel.

Person and number is marked on the verb, almost always as a prefix; only first person singular is marked as a suffix

The person /number markers on the verb indicate argument roles: subject, direct object, indirect object / dative case, benefactor etc.

Any argument may appear OR if it is clear from context it is possible to leave person / number argument markers out altogether;

The person / number markers that are suffixes come first, then follow the suffixes that indicate tense / aspect etc. ;

If a verb has multiple arguments that are all marked as prefixes, the Subject prefix comes first, then the Object prefix etc.;

The person / number argument markers that are used as affixes on verb stems are also affixed on adjectives to make a predicate (complement)

The table below breaks down the person / number argument markers used in this problem: hyphens indicate whether a prefix or suffix

Person:	Class 1 argument (like Subject / Nominative):	Class 2 argument (like Object / Accusative & inalienable possession):	Class 3 argument (like Indirect Object / Dative & alienable possession):
I	<i>-li</i>	<i>sa-</i>	<i>am-</i>

			<i>sam-</i> (when preceded by subject or object prefixes)
you (singular)	<i>ish-</i> <i>-is</i> (alternative if stem or following prefix begins with a sibilant)	<i>chi-</i>	<i>chim-</i>
he/she/it	unmarked	unmarked	not used in problem.

A note on tense & aspect marking on the verb:

The verb takes the suffix *-tok* in what equates to simple past tense (e.g. “I sang”) – a completed action that is now in the past;

The verb takes suffix *-h* to indicate a predicate which implies a tense that can be translated into English as either a present perfect tense (as in “I have sung”) or a progressive present tense (as in “I am singing”) – an action that is still having an influence on the present / now, and that is taking up some time;

There is no verb “to be” (the *-h* suffix on the adjective indicates that it is used predicatively as in sentence 4, where the adjective “niya” becomes “niyah”);